

壹、前言

統計分析方法在社會科學調查中廣為應用。在傳統的敘述統計中，平均數、眾數、中位數……等，常被做為分析調查的工具，此類分析結果都屬單一表徵，然而，在生活當中許多思維並非以單一面向而能言之。例如：人對天氣的冷熱感，什麼是「冷」？什麼是「熱」？假設用傳統方式問卷調查「你認為天氣達多少溫度就是熱？」經傳統統計方式取問卷之平均數來分析，若結果為23度。我們是否可以定只要氣溫高達23度就稱為「熱」？低於23度便稱為「冷」呢？另外，就大學畢業生每月薪水所得滿意度，若採用傳統方式做問卷調查「你認為大學畢業生每月薪水應是多少元？」可能有人答2.4萬到2.5萬、有人答2.3到2.6萬……等。那每人答的都是一區間，若以傳統方式統計則難以計算出「大學畢業生每月薪水應是多少？」因為在人類社會生活中，許多事物難以用二元方式來論定，所以利用模糊統計的方式來分析事物，更能呈現人類真實的感受，更能貼近事物的真實面。模糊理論自1965年發展至今，已漸漸從模糊數學理論研究及工程界之應用，推廣到社會科學界的傳統統計推論與方法論中。如 Lowen（1990）、Ruspini（1991）、Dubois 和 Prade（1991）、Tseng和Klein（1992）分別

提出對理念概似（approximate reasoning）的計量方法。在社會科學計量方法發展與應用亦有詳細的實例，如黃仁德與吳柏林（1995）提出臺灣短期貨幣需求函數穩定性的模糊統計檢定方法；吳柏林與楊文山（1997）對模糊統計在社會調查分析的應用；吳柏林與曾能芳（1998）考慮以模糊迴歸參數估計於景氣對策信號之迴歸模型分析，這些研究均獲得不錯的結果。

在傳統教育測驗統計上，試題難易程度都以答對率或得分率的方式來評定。用來表示每個試題的難易程度，一般使用答對百分比法來表示。答對百分比法，即是計算全體受試者中，答對的比例有多少。難度的值愈高，代表答對的人數愈多，也就是題目愈簡單。難度計算公式如下：

$$P = \frac{R}{N}$$

其中R為答對人數，N為作答學生總人數。答對率或得分率高代表學生通過率高，因此推論試題較容易；反之，答對率或得分率低則代表學生通過率低，因此推論試題較難。另一種常用來估算試題難度的方式採試題反應理論，試題反應理論（Item Response Theory, IRT）為近年進行試題分析的主要理論之一，此理論以個別試題的觀點，來對學生所得到的測驗分數進行解釋，而受試者在試題上答對與否，與其自身所具

備的能力或特質息息相關，而這種關係則透過連續遞增的數學函式來表示（余民寧，2009），透過試題反應理論的對數型模式，可以計算試題的難度 b 值。在試題反應理論中，單參數為IRT模組中最簡單，且為應用最廣的一種。對單參數的對數模型來說，唯一考量的是試題的難度。在此模組中假設鑑別度對每一個題目而言都是相同的，且猜測度都是不存在的。單參數的數學公式可以表達如下：

$$P_i(\theta) = P_i(X_i = 1 | \theta) = \frac{1}{1 + e^{-(\theta - b_i)}}$$

其中 $P_i(\theta)$ 代表某一位受試者，受試者的能力值為 θ ，而其答對第 i 題題目的機率。這個式子也可以用 $P_i(X_i = 1 | \theta)$ 來表示，代表當受試者的能力值固定為 θ ，他答對第 i 題題目的機率。至於 b_i 則是代表試題的難度。由這個公式可知，在單向度的單參數對數模型中，難度值是唯一會影響受試者答對題目機率的試題參數值。IRT的難度和傳統把通過率當作難度指標的意義是截然不同的。在IRT的理論中，難度愈高，代表題目愈難，而傳統把通過率當作難度指標，則是數值愈高，代表答對人數愈多，題目愈簡單。傳統的難度值容易受到抽取樣本的影響。若是剛好所選擇的受試者程度都很好，通過率就會高，就顯得題目很容易，若對相同的題目，給予一群程度比較差的學生來

考，那這題通過率就會低，題目變得很難。傳統的難度值，具有這種依賴抽樣的問題。而在IRT的難度值估計中卻可透過數學模組的轉換，使難度估計不會受到樣本的影響。

不管是建基於傳統試題理論的答對率，或是現代理論的難度，都需要有實徵數據。然而，對於一些高風險考試，如大學入學考試、國中基本學力測驗、或是高普考，題目一旦預試就有洩漏的危險，往往只能單靠闖場內專家的主觀認知判斷評定試題的難度。然而，常常這樣的主觀判斷會與實際的作答反應落差很大。例如，專家可能覺得某試題真的比較複雜而難以計算，但忽略大部分學生熟於此類型題目，造成答對率或得分率偏高，因此，本研究提出在這種情境下如何集結眾人智慧，在無法獲得學生作答反應實徵數據的情形下，將人們對試題的感受程度納入評判項目，並用以估算試題難度。誠如 George Edward Pelham Box所言：「All models are wrong but some are useful」，即使目前已有發展完善的古典與現代測驗理論，但單純以答對率或數學模型來評定試題的難易程度，缺少人們真正的感受程度。若能將人們對試題的感受程度納入，可以提供教師在設計試題時難度面向的另一項參考。

本研究以模糊軟計算方法代替傳統二元劃分方式，從大學學科能力測驗